

1. Breath alcohol tests in Ames, Iowa, USA

Ames, Iowa, USA is the home of Iowa State University, a land grant university with over 36,000 students. By comparison, the city of Ames, Iowa, itself only has about 65,000 residents. As with any other college town, Ames has had its fair share of alcohol-related incidents. (For example, Google 'VEISHEA riots 2014'.) We will take a look at some breath alcohol test data from Ames that is published by the State of Iowa.



The data file 'breath\alcohol\ames.csv' contains 1,556 readings from breath alcohol tests administered by the Ames and Iowa State University Police Departments from January 2013 to December 2017. The columns in this data set are year, month, day, hour, location, gender, Res1, Res2.

```
In [204]: # Load the packages
library(dplyr)
library(readr)
library(ggplot2)

# Read the data into your workspace
ba_data <- read_csv("datasets/breath_alcohol_ames.csv")

# Quickly inspect the data
head(ba_data)

# Obtain counts for each year
ba_year <- ba_data%>%
count(year, sort =T)

ba_year
```

Parsed with column specification:

```
cols(
  year = col_integer(),
  month = col_integer(),
  day = col_integer(),
  hour = col_integer(),
  location = col_character(),
  gender = col_character(),
  Res1 = col_double(),
  Res2 = col_double()
)
```

A tibble: 6 x 8

year	month	day	hour	location	gender	Res1	Res2
<int>	<int>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
2017	12	17	1	Ames PD	M	0.046	0.046
2017	12	14	3	ISU PD	F	0.121	0.120
2017	12	10	5	ISU PD	F	0.068	0.067
2017	12	10	3	ISU PD	F	0.077	0.077
2017	12	9	2	ISU PD	M	0.085	0.084
2017	12	9	1	Ames PD	M	0.160	0.161

A tibble: 5 x 2

year	n
<int>	<int>
2013	397
2014	344
2015	286
2016	269
2017	260

2. What is the busiest police department in Ames?

There are two police departments in the data set: the Iowa State University Police Department and the Ames Police Department. Which one administers more breathalyzer tests?

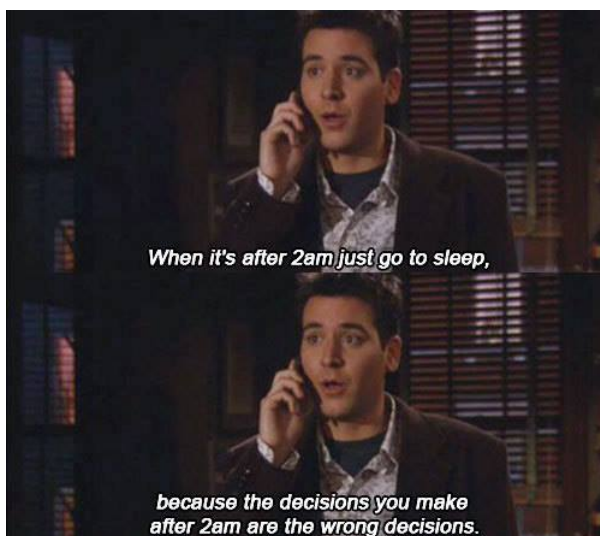
```
In [206]: # Count the totals for each department
pds <- ba_data%>%
count(location, sort = T)

head(pds)
```

A tibble: 2 x 2

location	n
<chr>	<int>
ISU PD	940
Ames PD	616

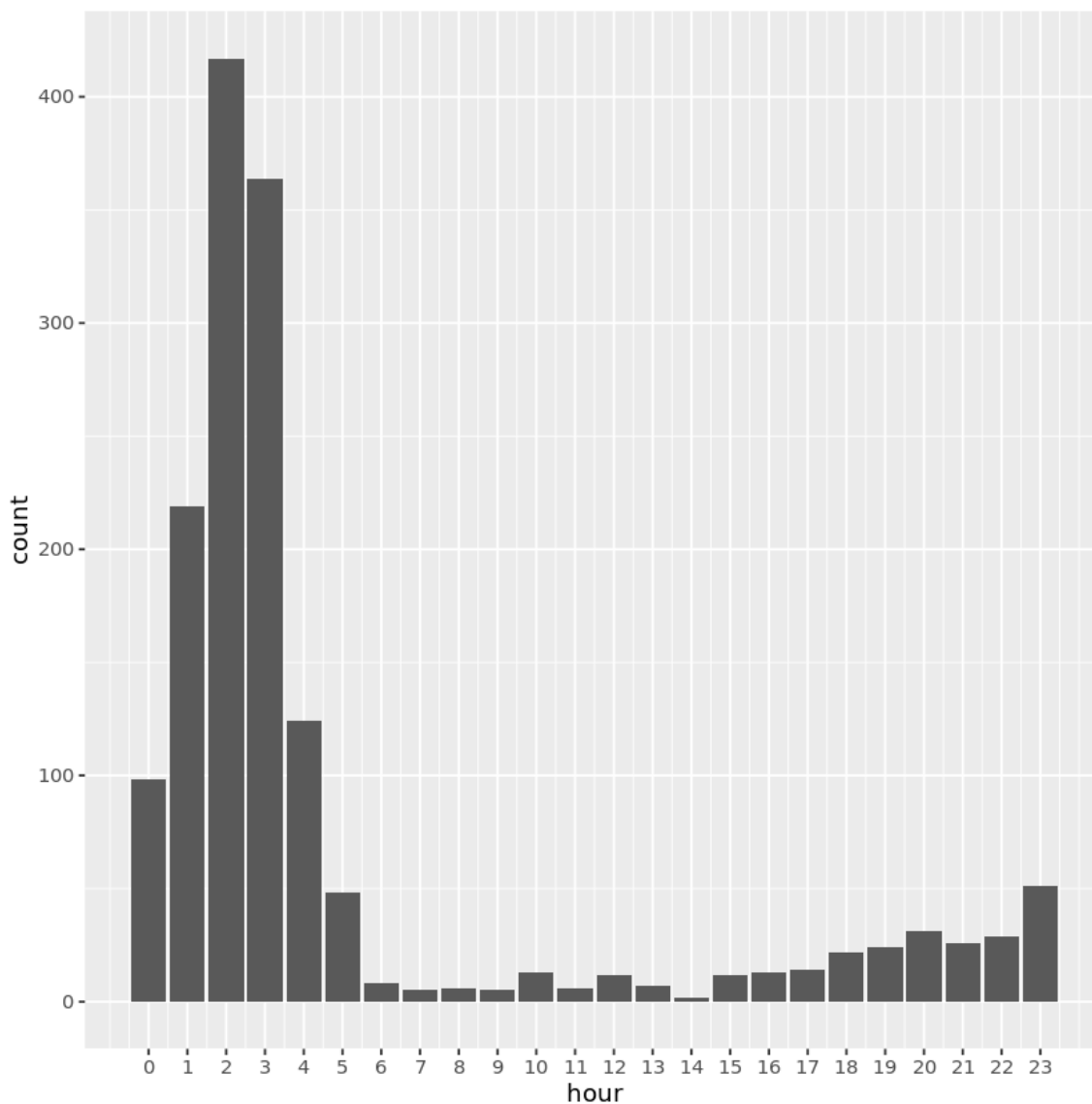
3. Nothing Good Happens after 2am



We all know that "nothing good happens after 2am." Thus, there are inevitably some times of the day when breath alcohol tests, especially in a college town like Ames, are most and least common. Which hours of the day have the most and least breathalyzer tests?

```
In [208]: # Count by hour and arrange by descending frequency
hourly <- ba_data%>%
  count(hour)%>%
  arrange(desc(n))

# Use a geom_bar
ggplot(hourly, aes(x = hour, weight = n)) + geom_bar() +
  scale_x_continuous(breaks = 0:23)
```



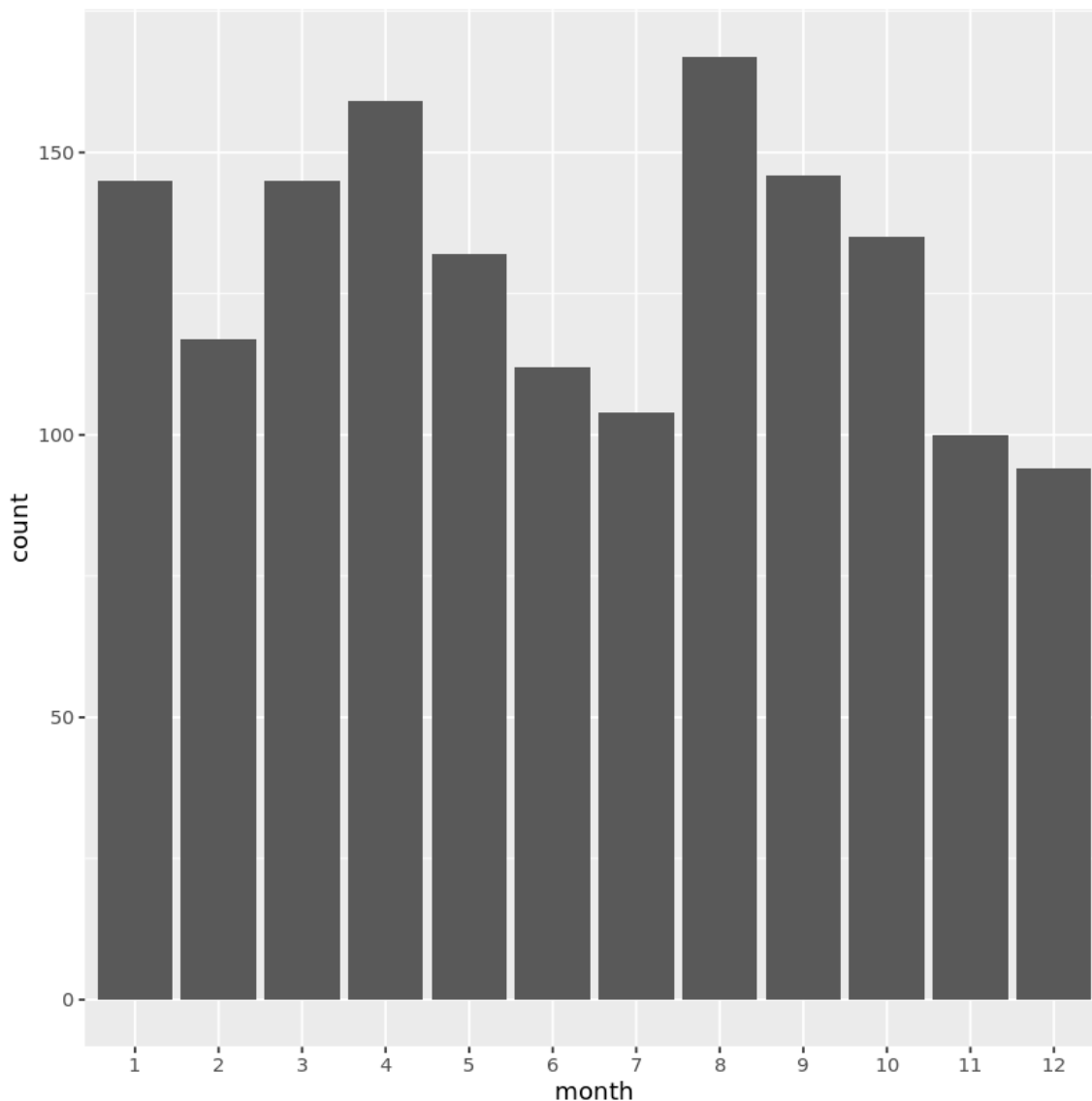
4. Breathalyzer tests by month

Now that we have discovered which time of day is most common for breath alcohol tests, we will determine which time of the year has the most breathalyzer tests. Which month will have the most recorded tests?

```
In [210]: # Count by month and arrange by descending frequency
monthly <- ba_data%>%
count(month, sort = T)

# Make month a factor
monthly$month <- as.factor(monthly$month)

# Use a geom_bar
ggplot(monthly, aes(x=month, weight=n)) + geom_bar()
```



5. COLLEGE

When we think of (binge) drinking in college towns in America, we usually think of something like this image at the left. And so, one might suspect that breath alcohol tests are given to men more often than women and that men drink more than women.



```
In [212]: # Count by gender
ba_data %>%
count(gender)

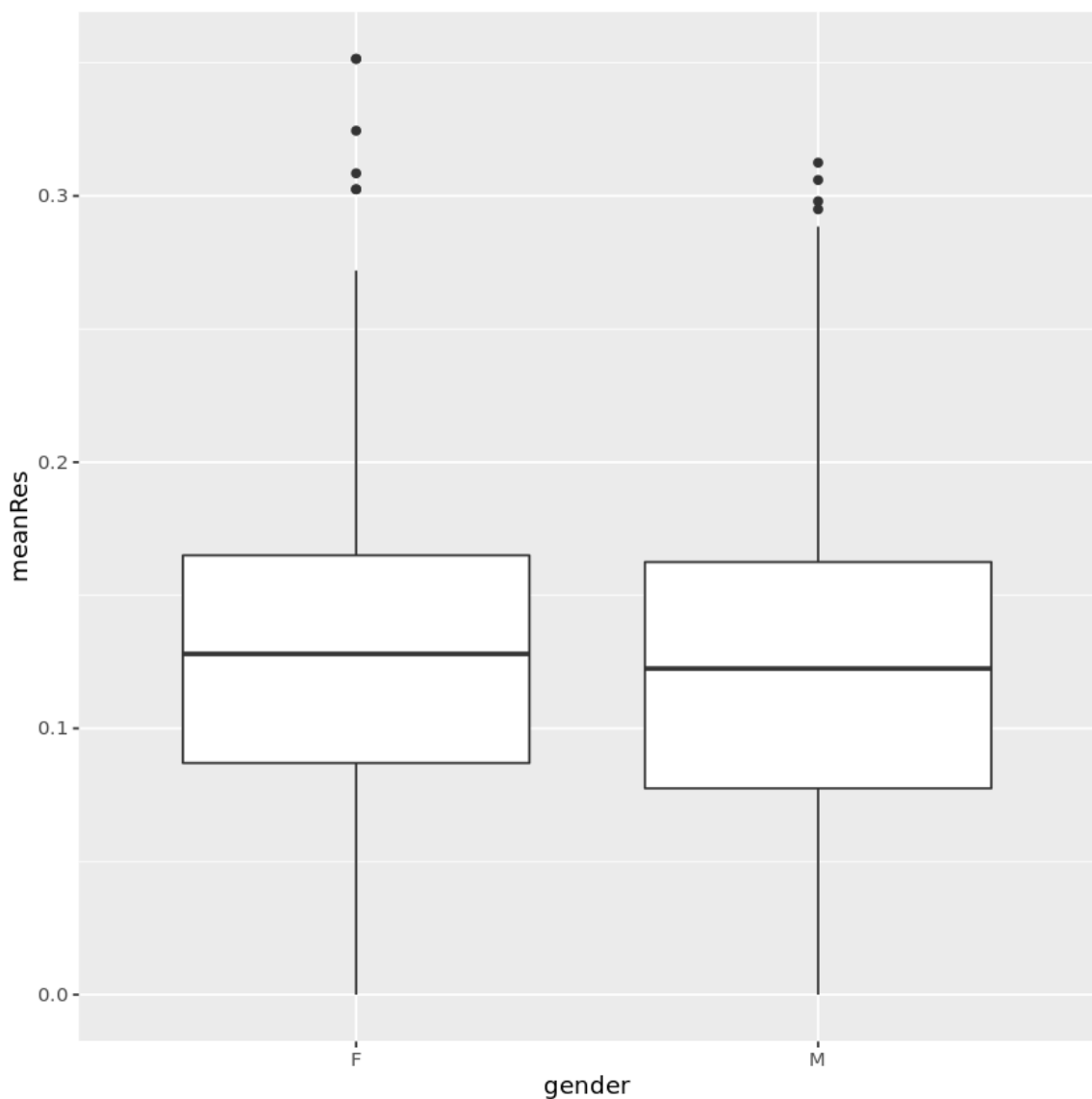
# Create a dataset with no NAs in gender
clean_gender <- ba_data%>%
filter(!is.na(gender))

# Create a mean test result variable and save as mean_bas
mean_bas <- clean_gender%>%
mutate(meanRes = (Res1 + Res2)/2)

# Create side-by-side boxplots to compare the mean blood alcohol levels of m
ggplot(mean_bas, aes(gender, meanRes)) + geom_boxplot()
```

A tibble: 3 x 2

gender	n
<chr>	<int>
F	425
M	1102
NA	29



6. Above the legal limit

In the USA, it is illegal to drive with a blood alcohol concentration (BAC) above 0.08%. This is the case for all 50 states (<https://www.dmv.org/automotive-law/dui.php>). Assuming everyone tested in our data was driving (though we have no way of knowing this from the data), if either of the results (Res1, Res2) are above 0.08, the person would be charged with DUI (driving under the influence).

```
In [214]: # Filter the data
duis <- ba_data%>%
filter(Res1 > 0.08 | Res2 > 0.08)

# Proportion of tests that would have resulted in a DUI
p_dui <- nrow(duis)/nrow(ba_data)
p_dui
```

0.744858611825193

7. Breathalyzer tests: is there a pattern over time?

We previously saw that 2 a.m. is the most common time of day for breathalyzer tests to be administered, and August is the most common month of the year for breathalyzer tests. Now, we look at the weeks in the year over time. We briefly use the lubridate (<http://lubridate.tidyverse.org/>), package for a bit of date-time manipulation.

```
In [216]: library(lubridate)

# Create date variable using paste() and ymd()
ba_data <- ba_data %>% mutate(date = ymd(paste(year, month, day)))

# Create a week variable using week()
ba_data <- ba_data %>% mutate(week = week(date))

head(ba_data)
```

A tibble: 6 x 10

year	month	day	hour	location	gender	Res1	Res2	date	week
<int>	<int>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>	<date>	<dbl>
2017	12	17	1	Ames PD	M	0.046	0.046	2017-12-17	51
2017	12	14	3	ISU PD	F	0.121	0.120	2017-12-14	50
2017	12	10	5	ISU PD	F	0.068	0.067	2017-12-10	50
2017	12	10	3	ISU PD	F	0.077	0.077	2017-12-10	50
2017	12	9	2	ISU PD	M	0.085	0.084	2017-12-09	49
2017	12	9	1	Ames PD	M	0.160	0.161	2017-12-09	49

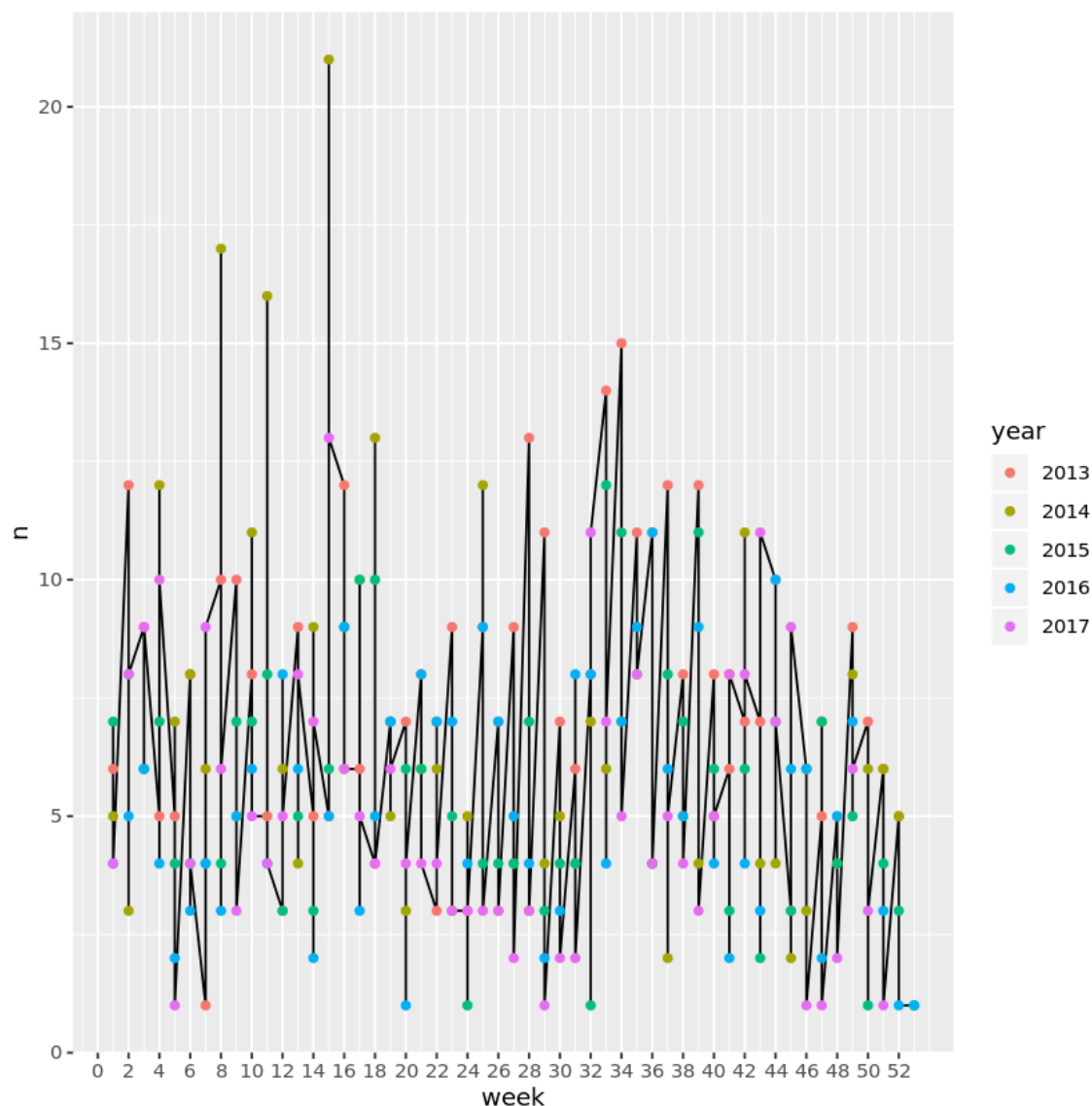
8. Looking at timelines

How do the weeks differ over time? One of the most common data visualizations is the time series, a line tracking the changes in a variable over time. We will use the new week variable to look at test frequency over time. We end with a time series plot showing frequency of breathalyzer tests by week in year, with one line for each year.

```
In [218]: # Create the weekly data set
weekly <- ba_data%>%
count(week,year)

# Make year a factor
weekly <- weekly %>%
mutate(year = as.factor(year))

# Create the time series plot with one line for each year
ggplot(weekly, aes(x = week, y = n)) +
  geom_line() +
  geom_point(aes(color = year)) +
  scale_x_continuous(breaks = seq(0,52,2))
```



9. The end of VEISHEA

From [Wikipedia \(https://en.wikipedia.org/wiki/VEISHEA\)](https://en.wikipedia.org/wiki/VEISHEA): "VEISHEA was an annual week-long celebration held each spring on the campus of Iowa State University in Ames, Iowa. The celebration featured an annual parade and many open-house demonstrations of the university facilities and departments. Campus organizations exhibited products, technologies, and held fundraisers for various charity groups. In addition, VEISHEA brought speakers, lecturers, and entertainers to Iowa State. [...] VEISHEA was the largest student-run festival in the nation, bringing in tens of thousands of visitors to the campus each year."

This over 90-year tradition in Ames was terminated permanently (<https://www.news.iastate.edu/news/2014/08/07/veisheaend>) after riots in 2014 (<https://www.desmoinesregister.com/story/news/crime-and-courts/2014/04/09/veishea-ames-car-tipping/7495935/>), where drunk celebrators flipped over multiple vehicles and tore light poles down. This was not the first incidence of violence and severe property damage in VEISHEA's history. Did former President Leath make the right decision?

```
In [220]: # Run this code to create the plot
ggplot() +
  geom_point(data = weekly, aes(x = week, y = n, color = year)) +
  geom_line(data = weekly, aes(x = week, y = n, color = year)) + # included
  geom_segment(data = NULL, arrow = arrow(angle = 20, length = unit(0.1, "inches"),
    ends = "last", type = "closed"),
    aes(x = c(20,20), xend = c(15.5,16), y = c(21, 20), yend = c(
  geom_text(data = NULL, aes(x = 23, y = 20.5, label = "VEISHEA Weeks"), size = 12) +
  scale_x_continuous(breaks = seq(0,52,2))

# Make a decision about VEISHEA. TRUE or FALSE?
cancelling_VEISHEA_was_right <- TRUE
```

